Social Meets Molecular: Combining Phylogenetic and Latent Class Analyses to Understand HIV-1 Transmission in Switzerland

Avila, Dorita ; Keiser, Olivia ; Egger, Matthias ; Kouyos, Roger ; Böni, Juerg ; Yerly, Sabine ; Klimkait, Thomas ; Vernazza, Pietro L ; Aubert, Vincent ; Rauch, Andri ; Bonhoeffer, Sebastian ; Guenthard, Huldrych F ; Stadler, Tanja ; Spycher, Ben D

Abstract: Switzerland has a complex human immunodeficiency virus (HIV) epidemic involving several populations. We examined transmission of HIV type 1 (HIV-1) in a national cohort study. Latent class analysis was used to identify socioeconomic and behavioral groups among 6,027 patients enrolled in the Swiss HIV Cohort Study between 2000 and 2011. Phylogenetic analysis of sequence data, available for 4,013 patients, was used to identify transmission clusters. Concordance between sociobehavioral groups and transmission clusters was assessed in correlation and multiple correspondence analyses. A total of 2,696 patients were infected with subtype B, 203 with subtype C, 196 with subtype A, and 733 with recombinant subtypes (mainly CRF02$_A$G and CRF01$_A$E). Latent class analysis identified 8 patient groups. Most transmission clusters of subtype B clusters with groups 4 and 5. Combining approaches from social and molecular epidemiology can provide insights into HIV-1 transmission and inform the design of prevention strategies.

DOI: https://doi.org/10.1093/aje/kwu076

Social Meets Molecular: Combining Phylogenetic and Latent Class Analyses to Understand HIV-1 Transmission in Switzerland

Dorita Avila, Olivia Keiser, Matthias Egger, Roger Kouyos, Jürg Böni, Sabine Yerly, Thomas Klimkait, Pietro L. Vernazza, Vincent Aubert, Andri Rauch, Sebastian Bonhoeffer, Huldrych F. Günthard, Tanja Stadler, Ben D. Spycher and the Swiss HIV Cohort Study

Correspondence:

Dr. Ben Spycher

Postal address: Institute of Social & Preventive Medicine (ISPM), University of Bern, Finkenhubelweg 11, CH-3012 Bern, Switzerland

Email: bspycher@ispm.unibe.ch

Tel: +41 31 631 5697

Fax: +41 31 631 3520

Acknowledgements

Author affiliations:

1) Institute of Social & Preventive Medicine (ISPM), University of Bern, Switzerland (Dorita Avila, Olivia Keiser, Matthias Egger, Ben D. Spycher)

2) Institute for Integrative Biology, ETH Zürich, Zürich, Switzerland (Dorita Avila, Sebastian Bonhoeffer)

3) Centre for Infectious Disease Epidemiology & Research, University of Cape Town, Cape Town, South Africa (Matthias Egger)

4) Institute of Medical Virology, Swiss National Center for Retroviruses, University of Zürich, Zürich, Switzerland (Jürg Böni)

5) Laboratory of Virology, Department of Genetics and Laboratory Medicine, University Hospitals of Geneva and Faculty of Medicine, University of Geneva, Geneva, Switzerland (Sabine Yerly)

6) Department Biomedicine - Petersplatz Building, University of Basel, Basel, Switzerland (Thomas Klimkait)

7) Division of Infectious Diseases, Cantonal Hospital St. Gallen, St. Gallen, Switzerland (Pietro L. Vernazza)

8) Division of Immunology and Allergy, Department of Laboratory Medicine, Lausanne University hospital, Lausanne, Switzerland (Vincent Aubert)

9) Department of Infectious Diseases, Bern University Hospital and University of Bern, Switzerland (Andri Rauch)

10) Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zürich, University of Zürich, Zürich, Switzerland (Huldrych F. Günthard, Roger Kouyos)

11) Department of Biosystems Science and Engineering, ETH Zürich, Zürich, Switzerland (Tanja Stadler)

Statement about authors contributing equally to the work:

Tanja Stadler and Ben Spycher contributed equally to this work

Grants and/or financial support

This work was funded in the framework of the Swiss HIV Cohort Study, supported by the Swiss National Science Foundation (SNF grant #33CS30-134277, 33CS30-148522, and PZ00P3-136820 (to TS), and SHCS projects #470, 528, 569, 687), the

SHCS Research Foundation, the Swiss National Science Foundation (grant #324730-130865, to HFG), the European Community's Seventh Framework Program (grant FP7/ 2007–2013), under the Collaborative HIV and Anti-HIV Drug Resistance Network (CHAIN; grant 223131, to HFG), the Yvonne-Jacob Foundation (to HFG), and by a further research grant of the Union Bank of Switzerland, in the name of an anonymous donor to HFG, an unrestricted research grant from Gilead, Switzerland to the SHCS research foundation, and by the University of Zurich's Clinical research Priority Program (CRPP) "Viral infectious diseases: Zurich Primary HIV Infection Study" (to HFG). O Keiser, R Kouyos, and B Spycher were supported by Swiss National Science Foundation fellowships (grants 32333B_131629/1, PZ00P3-142411 and PZ00P3_147987/1 respectively). No funding bodies had any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Thank-you's:

We thank the patients who participate in the SHCS; the physicians and study nurses for excellent patient care; the resistance laboratories for high-quality genotypic drug resistance testing; SmartGene, Zug, Switzerland, for technical support; Brigitte Remy, Dr. Martin Rickenbach, Dr. Franziska Schoeni-Affolter, and Yannick Vallet from the SHCS Data Center in Lausanne for data management; and Daniele Perraudin and Mirjam Minichiello for administrative assistance.

Members of the Swiss HIV Cohort Study:

Aubert V, Barth J, Battegay M, Bernasconi E, Böni J, Bucher HC, Burton-Jeangros C, Calmy A, Cavassini M, Egger M, Elzi L, Fehr J, Fellay J, Furrer H (Chairman of the Clinical and Laboratory Committee), Fux CA, Gorgievski M, Günthard H (President of the SHCS), Haerry D (deputy of "Positive Council"), Hasse B, Hirsch

HH, Hösli I, Kahlert C, Kaiser L, Keiser O, Klimkait T, Kouyos R, Kovari H, Ledergerber B, Martinetti G, Martinez de Tejada B, Metzner K, Müller N, Nadal D, Pantaleo G, Rauch A (Chairman of the Scientific Board), Regenass S, Rickenbach M (Head of Data Center), Rudin C (Chairman of the Mother & Child Substudy), Schöni-Affolter F, Schmid P, Schultze D, Schüpbach J, Speck R, Staehelin C, Tarr P, Telenti A, Trkola A, Vernazza P, Weber R, Yerly S.

Conflicts of Interest:

The authors have declared that no conflicts of interest exist.

Abbreviations

AIDS	Acquired Immune Deficiency Syndrome
HIV	Human immunodeficiency virus
IDU	Injection drug use
LCA	Latent class analysis
ML	Maximum likelihood
SEP	Socioeconomic position
SHCS	Swiss HIV Cohort Study

Running Head: HIV-1 Transmission in Switzerland

Word counts: Abstract 200 words, main text 3724 words, 1 box, 2 tables, 3 figures, 1 Web Appendix, 4 Web Tables, 5 Web Figures, 27 references.

Abstract

Switzerland has a complex human immunodeficiency virus (HIV) epidemic involving several populations. We examined HIV-1 transmission in a national cohort study. Latent class analysis was used to identify socioeconomic and behavioral groups among 6027 patients enrolled in the Swiss HIV Cohort Study between 2000 and 2011. Phylogenetic analysis of sequence data, available for 4031 patients, was used to identify transmission clusters. Concordance between socio-behavioral groups and transmission clusters was assessed in correlation and multiple correspondence analyses. 2697 patients were infected with subtype B, 203 with C, 196 with A and 733 with recombinant subtypes (mainly CRF02_AG, CRF01_AE). Latent class analysis identified eight patient groups. Most transmission clusters of subtype B were shared between groups of gay men (groups 1-3) or between the heterosexual groups 'heterosexual people of lower socioeconomic position (group 4) and 'injection drug users'(8). Clusters linking homosexual and heterosexual groups were associated with 'older heterosexual and gay people on welfare'(5). 'Migrant women in heterosexual partnerships'(6) and 'heterosexual migrants on welfare'(7) shared non-B clusters with 'heterosexual people of lower socio-economic position'(4) and 'older heterosexual and gay people on welfare'(5). Combining approaches from social and molecular epidemiology can provide insights into HIV-1 transmission and inform the design of prevention strategies.

Keywords: HIV-1 transmission, sexual orientation, injection drug use, socio-economic position, latent class analysis, phylogenetics, Switzerland

Switzerland has been more severely affected by the HIV type 1 (HIV-1) epidemic, as documented by a higher adult HIV-1 prevalence (0.4% in 2009) compared to other countries in Western Europe, for example Germany (0.1%) or The Netherlands (0.2%) (1). HIV-1 is the most common strain of the virus while the other major strain, HIV-2, is rarely seen outside of Africa. As is typical for Western Europe, the main risk groups include gay men, injection drug users and people acquiring HIV-1 through heterosexual intercourse. Whereas infections due to sharing of contaminated injecting equipment have been reduced considerably in Switzerland (1), the number of new infections in gay men has increased since 2000. This is probably due to continuing high risk sexual behavior, particularly by those unaware of their infection status (2). Furthermore, infections acquired abroad, either during travel or due to immigration from sub-Saharan Africa or Asia have become important, with an increasing proportion of non-B subtypes among individuals diagnosed in recent years (3, 4).

The rapid evolution of HIV-1 makes it possible to infer epidemiological patterns from sequence data (5, 6). In the Swiss HIV Cohort Study (SHCS), a nationwide cohort of people living with HIV/Aids in Switzerland (7), phylogenetic analyses found that the majority of clusters representing on-going transmission in Switzerland were linked to gay men (8). The previous studies (4, 8) focused on the main HIV transmission groups but these consist of heterogeneous groups of patients, including young and older gay men, migrants, people of higher or lower socio-economic position (SEP), or current and former injection drug users. An alternative approach to analyzing differences between transmission groups is to examine how relevant socio-behavioral characteristics cluster across patients. Originally developed in the social sciences, latent class analysis (LCA) (9) is increasingly used for this

purpose in medical research. The method assumes that the population is composed of distinct subpopulations (latent classes), which are not directly observed but inferred from the observed characteristics of individuals. In the SHCS, LCA was recently used to identify socio-behavioral groups and to examine outcomes of antiretroviral therapy across these groups (10).

In the present study, we sought to combine social and molecular epidemiology to investigate likely transmission patterns of HIV-1, based on phylogenetic analyses of *pol* sequences, between socio-behavioral groups identified in latent class analysis.

MATERIALS AND METHODS

Study population

Established in 1988, the Swiss HIV Cohort Study (SHCS) is an on-going nationwide, prospective cohort study, which includes information on over 15,000 patients (7). A comparison of cohort data with official AIDS notifications indicated that about 70% of all individuals living with HIV/AIDS in Switzerland participate in the SHCS (11). A recent comparison with drug sales data showed that 75% of antiretroviral drug prescriptions in Switzerland can be attributed to study participants (7). Sequence data from routine viral load tests and systematic retrospective sequencing from the SHCS sample repository are available for about 70% of the enrolled patients. Starting from the year 2000, a questionnaire on socio-demographic and behavioral data has been completed by the treating physician at the time of enrolment and during semi-annual visits. All participants provided informed consent.

Identification of socio-behavioral groups

We used latent class analysis (LCA) (9, 12) to identify groups of patients with similar socio-demographic and behavioral characteristics. This analysis was an update of a previously published LCA (10). We included all individuals who enrolled between 2000 and 2011. LCA included the following categorical variables at registration into the cohort: sex, age (categories <25, 25-34, 35-44, 45-54, ≥55 years), region of origin (Switzerland and North-West Europe, Southern Europe, Sub-Saharan Africa, Latin America, Asia and Eastern Europe), level of education (compulsory schooling, vocational training, higher education), occupation (self-employed, apprentice or trainee, higher management, middle or lower management level, employee,

housewife/homemaker) and main source of income (salaried work, support from family or partner; welfare benefits), sexual preference (heterosexual, bisexual, homosexual), sexual contacts (no partner, unprotected sex with stable partner, protected sex with stable partner, unprotected sex with occasional partner, protected sex with occasional partner), history of injection drug use (IDU) (never, ever, current) and alcohol consumption (severe, moderate, light according to World Health Organization (13)). These variables cover all demographic, socioeconomic and behavioral factors on which data is collected in the SHCS and which, in our view, might be associated with transmission. To characterize patient groups identified by LCA socio-economic position (SEP) of patients was defined as higher or lower, based on educational attainment and occupation.

LCA assumes that the population consists of subpopulations (latent classes) that differ in the distribution of the included variables, and that within latent classes these variables are independent. We used the Bayesian Information Criterion to select the number of classes, thus balancing parsimony and model fit (12). For each patient, we computed posterior probabilities of belonging to the different latent classes of a fitted model. We allocated individuals to the groups for which they had the highest posterior membership probability (12). We fitted the LCA models using Mplus version 6.1 (14). More details of the LCA are provided in the Web Appendix 1.

Phylogenetic analysis

All laboratories performing genotypic resistance testing in Switzerland participate in the SHCS drug resistance database (15). They sequence at least the full protease and part of the reverse transcriptase (RT) using in-house methods (16) or commercial

assays (ViroSeq Vs.1 PE Biosystems, Rotkreuz, Switzerland; ViroSeq Vs.2, Abbott AG, Baar, Switzerland; vircoTYPE HIV-1 Assay, Virco Lab, Mechelen, Belgium). For the phylogenetic analysis, we included all patients for whom sequence data was available. In case of multiple sequences, the oldest sequence was retained.

In order to identify domestic transmission clusters, we first split the data into the major subtypes. Within a subtype, we then randomly selected the same number of sequences from other countries using the Los Alamos Sequence Database (8, 17). Viral sequences from domestic transmission chains are expected to cluster with SHCS sequences rather than with Los Alamos sequences, while sequences from infections acquired abroad will not show such clustering. Antiretroviral treatment may lead to convergent evolution, thus distorting phylogenetic analyses. We therefore removed all major amino-acid positions associated with antiretroviral drug resistance following the International AIDS Society (IAS)–USA guidelines (18).

We inferred the maximum-likelihood (ML) tree per major subtype with a GTRCAT model implemented in RaxML (19) and used 100 bootstrap trees to define the strength of support for nodes in the ML tree. A subtype K Los Alamos sequence was used as outgroup, i.e. a reference group related to all sequences but whose relationship is expected to be less close than any of the other sequences. Trees were plotted with FigTree (http://tree.bio.ed.ac.uk/software/figtree/). For every subtype, the ML tree was explored using the *ape* package in the statistical software R (20). We inferred transmission clusters as described in Hue et al. (6) and as in previous studies of the SHCS (4, 8). SHCS clusters were defined as clusters with >70% bootstrap support and >70% of individuals from the SHCS.

Concordance of socio-behavioral groups and phylogenetic clusters

We cross-tabulated socio-behavioral groups (columns) against SHCS phylogenetic clusters (rows) and analyzed the table in two ways. First, we calculated pair-wise Pearson correlation coefficients between columns, i.e. the correlation between each pair socio-behavioral groups regarding the frequency with which they occur in the different clusters. A substantial correlation between two groups indicates that patients from these groups appear in the same phylogenetic clusters, suggesting between-group transmission. Coefficients up to 0.4 were considered to reflect weak correlations, coefficients between 0.4 and 0.7 moderate correlations and coefficients above 0.7 strong correlations.

Second, we used Multiple Correspondence Analysis to obtain a two-dimensional representation of columns and rows and their interdependencies (21). In the Multiple Correspondence Analysis plots, socio-behavioral groups lie close together if their members appear in the same phylogenetic clusters, while phylogenetic clusters are positioned close to each other if they have a similar composition of socio-behavioral groups (21). In the full-dimensional Multiple Correspondence Analysis representation, of which we only display the two main dimensions, phylogenetic clusters are positioned at the barycenter (center of gravity) of the socio-behavioral groups they consist of, with the socio-behavioral groups having weights equal to their proportion in the phylogenetic cluster. Analyses were done for all subtypes, subtype B and non-B subtypes.

RESULTS

Study population

All 6027 patients who enrolled in the SHCS from 2000 to 2011 were included in the LCA analyses. Among these 4013 patients (66.6%) had viral sequence data available and 3581 patients (59.4%) had eligible subtypes. Finally, 1662 patients (27.6%) were part of a SHCS cluster (Figure 1). Patients with viral sequences were similar to all patients enrolled in the SHCS with respect to age and gender, the history of IDU, self-reported sexual contacts, region of origin and main source of income (Table 1). The proportion of heterosexual patients was higher in those included in the phylogenetic analyses (54.8% versus 37.5% among all patients). As expected, patients included in SHCS clusters were more likely to be from Switzerland or North-West Europe compared to all patients. They were also more likely to be gay men and to report former or current IDU (Table 1).

Socio-behavioral groups

The LCA identified eight distinct socio-behavioral groups, which are described in the Appendix and Web Table 1. A comparison of fit statistics for models with 1-10 classes and model results for the selected 8-class model are given in Web Tables 2 and 3 respectively. Three of the eight groups were dominated by gay men but differed with respect to age, SEP, migrant status and sexual partnership patterns. The three groups are numbered 1 to 3 and described as ‘Gay and bisexual men of lower SEP’, ‘Gay and heterosexual men of higher SEP’ and ‘Young gay migrant men’. Groups 4 to 7 were dominated by people who acquired HIV heterosexually. These groups again differed in terms of age, SEP, migrant status and partnerships and, additionally,

regarding dependence on welfare benefits. The four groups are described as 'Heterosexual people of lower SEP', 'Older heterosexual and gay people on welfare benefits', 'Migrant women in heterosexual partnerships', and 'Heterosexual migrants on welfare benefits'. Group 8 consisted of people with former or current IDU ('Injection drug users'). The posterior probabilities of belonging to the different latent classes ranged from 0.78 among gay and heterosexual men of higher SEP to 0.98 among injection drug users.

Phylogenetic analysis

Of the 4013 patients with phylogenetic data, 2697 (67.2%) were infected with subtype B. Other major subtypes included CRF02_AG (272, 6.8%), CRF01_AE (214, 5.3%), C (203, 5.1%), and A (196, 4.9%). The remaining subtypes were poorly represented (<2%) and excluded (Figure 1). The ML phylogenetic trees for all major subtypes (A, B, C, CRF02_AG and CRF01_AE) are shown in Web Figures 1-5.

The prevalence of the different subtypes across the eight socio-behavioral groups is shown in Figure 2. Subtype B dominated in the three groups of gay men (87% to 94%). Injection drug users had the highest prevalence of subtype B (95%) whereas heterosexual migrants on welfare benefits had the lowest prevalence (11%). Subtypes of African origin (A, 02_AG, C) dominated the latter group (85%) and were also highly prevalent among migrant women in heterosexual partnerships (56%). The subtype of Asian origin (01_AE) was also prevalent among migrant women in partnerships (20%), older heterosexual and gay people on welfare benefits (9%) and heterosexual people of lower SEP (8%).

Phylogenetic clusters

We identified 195 SHCS clusters including a total of 1964 sequences, of which 1662 belonged to SHCS patients. There were 140 B subtype clusters and 55 clusters of non-B subtypes: A (11 clusters), C (13), 01_AE (14) and 02_AG (17). The mean number of SHCS patients in SHCS clusters was larger for B clusters than non-B clusters B: 10.0 (range 2-150), A: 6.2 (2-21), C: 3.1 (2-7), 01_AE: 6.9 (2-21), 02_AG: 3.6 (2-8). Thirteen B clusters had over 20 SHCS sequences.

Concordance of socio-behavioral groups and phylogenetic clusters

Table 2 shows Pearson correlation coefficients between socio-behavioral groups and phylogenetic clusters for all virus subtypes, B subtype and non-B subtypes (95% confidence intervals for these coefficients are reported in the Web Table 4). When including all subtypes moderate or strong correlations were found between the groups of gay men (groups 1-3, r=0.59 to r=0.77). Somewhat weaker correlations were observed between these groups and the group of hetero- and homosexual people on welfare benefits (group 5, r=0.34 to r=0.67). The latter group also correlated with the group of heterosexual people of lower SEP (group 4, r=0.75) and the injection drug users (group 8, r=0.65). Migrant women in partnerships (group 6) were moderately associated with the group of heterosexual people of lower SEP (group 4, r=0.43). Heterosexual migrants on welfare benefits (group 7) were only weakly associated with the other groups with predominantly heterosexual orientation (r=0.24 to r=0.35). Finally, injection drug users showed a strong association with the heterosexual people of lower SEP (r=0.91).

When restricting analyses to subtype B, correlations tended to get stronger. In particular correlations involving the groups of migrants and those involving injection

drug users became more prominent. Many correlations disappeared or weakened when restricting analyses to non B subtypes while one new association (r=0.53) emerged for non-B subtypes between migrant women in heterosexual partnerships and homosexual young migrant men (Table 2).

Geometric representation of data using Multiple Correspondence Analysis

Associations between socio-behavioral groups and SHCS clusters are shown geometrically in Figure 3 for all subtypes (part A), subtype B (part B) and non B subtypes (part C). In all analyses the first two axes explained over 50% of the association (as measured by the χ^2-metric) between socio-behavioral groups and phylogenetic clusters. In the analysis of all subtypes, the three groups dominated by homosexual men (groups 1-3) are located in close vicinity around a concentration of clusters with sequences almost entirely from these groups. To the bottom right are clusters containing predominately sequences from heterosexual people of lower SEP (group 4) and injection drug users (group 8). In between, in the vicinity of the group of hetero- and homosexual people on welfare benefits (group 5), there is another concentration of phylogenetic clusters containing sequences of both homosexual and heterosexual patient groups. Smaller clusters with sequences from the two heterosexual migrant groups are found towards the right upper corner (groups 6 and 7). Thus axis 1 distinguished men who have sex with men dominated phylogenetic clusters from heterosexual and IDU dominated clusters while axis 2 distinguished small clusters involving heterosexual migrant groups from the large clusters involving heterosexual Swiss dominated groups. Patterns were similar when restricting the analysis to B subtypes, with a reduced representation of migrant groups (middle panel).

The picture changed when focusing on non-B subtypes: the two heterosexual migrant groups (groups 6 and 7) and the group of heterosexual men and woman of lower SEP (group 4) were now closely associated, with many shared phylogenetic clusters (lower right panel). The group of hetero- and homosexual people on welfare benefits (group 5) were found in the vicinity. Furthermore, homosexual migrant men (group 3) and homo- and heterosexual men of higher SEP (group 2) were represented both in said clusters and in few clusters associated with injection drug users (group 8).

DISCUSSION

By combining socio-behavioral data with phylogenetic transmission clusters, we characterized the Swiss HIV epidemic in greater depth and identified possible routes of HIV transmission between groups. These groups were described not only by sexual orientation or IDU, but also by socio-economic and migrant status or main source of income. We found phylogenetic clusters that were shared between homosexual and heterosexual patient groups, suggesting that there is transmission between these sub-epidemics. In particular, the group of older heterosexual and gay people on welfare benefits might act as a bridge between homosexual and heterosexual clusters, whereas injection drug users were associated with heterosexual people of lower SEP. Domestic transmission of non-B subtypes happened primarily through heterosexual contact with migrants: migrant women in partnerships and single migrants on welfare shared phylogenetic clusters with heterosexual people of lower SEP and older hetero- and homosexual people supported by welfare.

Limitations

Much of the information used to define the socio-behavioral groups was self-reported and therefore potentially biased. For example, unprotected sex and sex between men may have been underreported and protected and heterosexual sex over reported. Reporting behavior may have differed by socio-economic and migrant status. The collection of some of the socio-demographic information commenced in 2000, which limited the size and period covered by the study, and some relevant information was missing, for example on commercial sex work.

Sequences were only available for about 70% of patients included in the latent class analysis, but patients with sequences were similar to all patients. Routine resistance testing in drug naïve patients was introduced from 2003 onwards. Patients who started antiretroviral therapy earlier and never failed did not have a resistance test. Furthermore patients presenting with AIDS-defining conditions often receive antiretroviral therapy in the hospital and are undetectable when they are enrolled in the SHCS. We assigned patients to socio-behavioral groups based on highest membership probabilities, thus ignoring membership uncertainty. However because mean membership probabilities for the most likely class were high (>0.76, Web Table 3) we do not think this had a major impact on our results. The phylogenetic analysis is limited by common caveats, such as incompleteness of transmission chains and sampling as well as temporal biases affecting the proportion of subtypes. The high national representativeness of the dataset makes such biases less likely. Bootstrap support in the best trees was low in many of the clusters: this could have resulted in loss of domestic clusters. The size of clusters was typically small for non-B subtypes, which limited the power to detect associations between socio-behavioral groups for these subtypes.

Finally, we interpreted co-occurrence in the same phylogenetic cluster as indicating increased likelihood of transmission between two individuals. We acknowledge that there might be assortative mixing among patients whose viral sequences are in the same phylogenetic cluster. For instance, a phylogenetic cluster with sequences from both gay men and heterosexual people is compatible both with frequent mixing between these groups and predominant mixing within groups, with few transmissions between groups.

Comparison with previous studies of HIV-1 transmission in Switzerland

Kouyos et al. previously investigated subtype B transmission in Switzerland using data from the SHCS and similar phylogenetic methods (8). As in the present study, they found that most domestic transmission clusters fell into one of two broad groups: those dominated by homosexual transmission among men and those dominated by transmission through heterosexual contact and IDU. Their analysis did not allow a more detailed characterization of the patient groups involved. For example, the analysis could not identify the different migrant groups associated with the men who have sex with men epidemic and the heterosexual epidemic. Von Wyl et al. found that the epidemic of non-B subtypes was primarily maintained through re-introduction from abroad, but that domestic heterosexual transmission also played a role (4). Their results are in line with the findings of the present study which additionally showed that this epidemic specifically involves Swiss or European people of lower SEP, migrant woman in stable partnerships, but also migrants without stable partnerships.

Possible mechanisms

Transmission between groups of gay men and different heterosexual groups, as indicated by our data, would likely occur through men who have sex both with men and women. All groups dominated by homosexual men included a minority of men reporting bisexual orientation (8.5% to 11.1%). Bisexual orientation was rarely reported among the other groups, with the exception of the older heterosexual and gay people on welfare (6.6% bisexual) and injection drug users (3.8% bisexual). Switzerland lacks a dedicated survey of sexual behavior, such as the British National Survey of Sexual Attitudes and Lifestyles (22), but based on other surveys (23), the

prevalence of bisexuality reported by some groups, for example people of lower SEP (0.1%), was lower than anticipated, and thus compatible with underreporting.

The largest clusters of heterosexual transmission were of the B-subtype and dominated by heterosexual people of lower SEP and injection drug users. Heterosexual people of lower SEP also played an important role in the transmission of non-B subtypes while injection drug users did not, suggesting that, among heterosexuals, those who engage with migrants are different from those in contact with injection drug users. It seems likely that there is more heterogeneity with respect to sexual mixing between predominantly heterosexual groups, which was not fully captured by our socio-behavioral classification.

The group of older heterosexual and gay men is of particular interest for several reasons. It was represented in both the heterosexual clusters involving injection drug users and the smaller clusters involving migrant groups but also in clusters dominated by men who have sex with men. This heterogeneous group of older HIV-1 positive people is rapidly growing in Switzerland and elsewhere (24) and includes both people ageing with the disease and later life-acquired infections. These men may perceive themselves to be less at risk for HIV (24) and may be socially more isolated than younger men and therefore more vulnerable (25). Some of their contacts with either young homosexual migrants or young migrant women may have involved commercial sex. The groups of lower SEP may be more vulnerable and more difficult to reach, and these groups also included many older people. Migrants are another fast growing segment of the Swiss HIV positive population. Although many infections are likely acquired abroad, our data show that different groups of gay men and heterosexual groups shared phylogenetic clusters with migrant groups,

suggesting that there is domestic propagation, perhaps to some extent driven by commercial sex and older men.

Conclusions

In this study, LCA enabled us to characterize the subgroups involved in the Swiss HIV epidemic in greater detail than in previous phylogenetic studies of the Swiss cohort (4, 8), while the phylogenetic analyses uncovered associations between socio-behavioral groups not previously documented. The combination of social science and epidemiological tools in this study represents a novel approach, which should help inform the development of strategies to prevent the spread of HIV in Switzerland. In particular, specific HIV prevention strategies for older men, bisexual men and migrants should be prioritized - strategies that respect the human rights of HIV positive people and avoid stigmatization. Indeed, qualitative research into the needs and preferences of the groups identified in this study to ensure that any targeted health promotion programs are not inadvertently stigmatizing is a logical next step (26).

References

1. Joint United Nations Programme on HIV/AIDS (UNAIDS). Global report: UNAIDS report on the global AIDS epidemic 2010. Geneva: UNAIDS, 2010.
2. van Sighem A, Vidondo B, Glass TR, et al. Resurgence of HIV infection among men who have sex with men in Switzerland: mathematical modelling study. *PloS one* 2012;7(9):e44819.
3. Yerly S, von Wyl V, Ledergerber B, et al. Transmission of HIV-1 drug resistance in Switzerland: a 10-year molecular epidemiology survey. *Aids* 2007;21(16):2223-2229.
4. von Wyl V, Kouyos RD, Yerly S, et al. The role of migration and domestic transmission in the spread of HIV-1 non-B subtypes in Switzerland. *The Journal of infectious diseases* 2011;204(7):1095-1103.
5. Leventhal GE, Kouyos R, Stadler T, et al. Inferring epidemic contact structure from phylogenetic trees. *PLoS computational biology* 2012;8(3):e1002413.
6. Hue S, Pillay D, Clewley JP, et al. Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. *Proceedings of the National Academy of Sciences of the United States of America* 2005;102(12):4425-4429.
7. Swiss H. I. V. Cohort Study, Schoeni-Affolter F, Ledergerber B, et al. Cohort profile: the Swiss HIV Cohort study. *International journal of epidemiology* 2010;39(5):1179-1189.
8. Kouyos RD, von Wyl V, Yerly S, et al. Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. *The Journal of infectious diseases* 2010;201(10):1488-1497.
9. Lazarsfeld PF, Henry NW. *Latent structure analysis*. New York,: Houghton; 1968.
10. Keiser O, Spycher B, Rauch A, et al. Outcomes of antiretroviral therapy in the Swiss HIV Cohort Study: latent class analysis. *AIDS and behavior* 2012;16(2):245-255.
11. Gebhardt M, Rickenbach M, Egger M. Impact of antiretroviral combination therapies on AIDS surveillance reports in Switzerland. Swiss HIV Cohort Study. *Aids* 1998;12(10):1195-1201.
12. McLachlan G, Peel D. *Finite mixture models*. New York: John Wiley & Sons; 2000.
13. World Health Organization. Comparative quantification of health risks - Global and regional burden of disease attribution to selected major risk factors. Geneva, 2004.
14. Muthén BO, Muthén LK. *Mplus [Computer software]*. Los Angeles, CA: Muthén & Muthén; 1998.
15. von Wyl V, Yerly S, Boni J, et al. Emergence of HIV-1 drug resistance in previously untreated patients initiating combination antiretroviral treatment: a comparison of different regimen types. *Archives of internal medicine* 2007;167(16):1782-1790.
16. Yerly S, Vora S, Rizzardi P, et al. Acute HIV infection: impact on the spread of HIV and transmission of drug resistance. *Aids* 2001;15(17):2287-2292.
17. Los Alamos HIV Sequence Database. Los Alamos National Laboratory. http://www.hiv.lanl.gov/. Accessed November 3, 2011.
18. Hirsch MS, Gunthard HF, Schapiro JM, et al. Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society-USA panel. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2008;47(2):266-285.
19. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 2006;22(21):2688-2690.
20. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2012.
21. Greenacre M. *Theory and Application of Correspondence Analysis,*. London: Academic Press; 1984.
22. Johnson AM, Mercer CH, Erens B, et al. Sexual behaviour in Britain: partnerships, practices, and HIV risk behaviours. *Lancet* 2001;358(9296):1835-1842.

23. Wang J, Hausermann M, Wydler H, et al. Suicidality and sexual orientation among men in Switzerland: findings from 3 probability surveys. *Journal of psychiatric research* 2012;46(8):980-986.
24. Sankar A, Nevedal A, Neufeld S, et al. What do we know about older adults and HIV? A review of social and behavioral literature. *AIDS care* 2011;23(10):1187-1207.
25. Slater LZ, Moneyham L, Vance DE, et al. Support, stigma, health, coping, and quality of life in older gay men with HIV. *The Journal of the Association of Nurses in AIDS Care : JANAC* 2013;24(1):38-49.
26. Valdiserri RO. HIV/AIDS stigma: an impediment to public health. *American journal of public health* 2002;92(3):341-342.

Appendix. Characteristics of the Eight Socio-Behavioural Groups Identified by Latent Class Analysis in the Swiss HIV Cohort Study.

1) Gay and bisexual men of lower SEP *(1478 participants, 24.5%)* Median age at enrolment in the cohort was 37 years (interquartile range [IQR]: 31–42 years) and most men were from Switzerland or Northern or Western Europe (85%). Most reported homosexual (85%) or bisexual (11%) orientation. Two thirds (67%) reported occasional partners and 34% reported unprotected sex with occasional partners. Most men had vocational training (77%) and worked as manual or non-manual employees (52%).
2) Gay and heterosexual men of higher SEP *(873 participants, 14.5%)* Median age was 43 years (IQR 37–49 years) and most were from Switzerland or Northern or Western Europe (85%). The men reported mainly homosexual (73%) or heterosexual (17%) orientation. Over half (54%) reported occasional partners, with 29% reporting unprotected sex with occasional partners. Almost all of these men had a University degree (97%) and most (66%) worked in middle or higher management or were self-employed.
3) Young gay migrant men *(316 participants, 5.2%)* Median age of these men was 30 years (IQR 26–34 years) and most (63%) were migrants from Latin America, Asia or Eastern Europe. Sexual preference was predominantly homosexual (83%), with almost half (49%) reporting stable partnerships and 21% reporting unprotected sex with occasional partners. The majority (67%) had compulsory schooling only or vocational training, with a substantial proportion still in training (17%).
4) Heterosexual people of lower SEP *(1089 participants, 18.1%)* Median age in this group was 39 years (IQR 33–44 years); about half were women (47%). About a quarter (25%) was from sub-Saharan Africa or Southern Europe. Sexual preference was almost exclusively heterosexual (99%). Some (11%) reported a history of injection drug use. Most had stable partners (65%); few (9%) reported unprotected sex with occasional partners. The majority (62%) had vocational training or compulsory schooling only, and worked as manual or non-manual employees (58%).
5) Older Heterosexual and gay people on welfare benefits *(443 participants, 7.4%)* Median age was 59 years (IQR 52–64 years), 24% were women and most (90%) were from Switzerland and

Northern and Western Europe. Sexual orientation was predominantly heterosexual (76%), with 23% indicating homosexual or bisexual preferences. Many (43%) reported having no partner, with some (12%) reporting unprotected sex with occasional partners. The majority (74%) depended on welfare benefits and was out of work.
6) Migrant women in heterosexual partnerships *(546 participants, 9.1%)* Median age was 31 years (IQR 27–36 years). Most women were from sub-Saharan Africa (50%) or Asia and Eastern Europe (24%). Most women were heterosexual (98%) and reported stable partnerships (87%), with only very few (2%) reporting unprotected sex with occasional partners. The majority had compulsory schooling only (60%) and relied on support from their partners.
7) Heterosexual migrants on welfare benefits *(578 participants, 9.6%)* The median age in this group was 31 years (IQR 27–36 years); 61% were women and most were from sub-Saharan Africa (92%). Sexual orientation was predominantly heterosexual (97%) with about half of group members (47%) reporting no stable partner and a minority (18%) reporting sex with occasional partners. The majority had compulsory schooling only (68%) and depended on welfare benefits (78%).
8) Injection drug users *(704 participants, 11.7%)* The median age was 36 years (31–41 years); 35% were women. The majority (60%) reported former injection drug use, with 32% reporting current use and the remaining reporting no use. This group was mainly from Switzerland or Northern and Western Europe (79%). Sexual orientation was predominantly heterosexual (94%), with 20% reporting sex with occasional partners. The main source of income was welfare benefits (84%).

See Web Appendix 1 and Web Table 1 for a detailed compilation of the variables used to define these groups.

Table 1. Characteristics of Patients Included in the Latent Class Analysis, the Phylogenetic Analysis and the Analysis of SHCS Clusters, 2000-2011.

	Latent class analysis (n=6027)		Phylogenetic analysis (n=4013)		analysis of SHCS clusters (n=1662)	
	No.	**%**	**No.**	**%**	**No.**	**%**
No. of women	1766	29.3	1104	27.5	327	19.7
Age[a]	37 (35-45)		37 (35-45)		37 (35-45)	
Region of origin						
Switzerland and North-Western Europe	3761	62.4	2578	64.2	1234	74.2
Southern Europe	461	7.6	398	9.9	135	8.1
Sub-Saharan Africa	977	16.2	596	14.9	90	5.4
Latin America	250	4.1	157	3.9	44	2.6
Asia, Eastern Europe	392	6.5	254	6.3	114	6.9
Other or unknown	186	3.1	120	3.0	45	2.7
Occupation						
Student or trainee	162	2.7	121	3.0	40	2.4
Employee	2569	42.6	1931	48.1	836	50.3
Middle or lower management	520	8.6	413	10.3	181	10.9
Higher management	209	3.5	157	3.9	66	4.0
Self-employed	599	9.9	447	11.1	179	10.8
Housewife/homemaker	230	3.8	154	3.8	41	2.5
Other or unknown	1738	28.8	790	19.7	319	19.2
Main source of income						
Salaried work	3554	59.0	2464	61.4	1095	65.9
Support from family or partner	656	10.9	399	9.9	117	7.0
Welfare benefits	1756	29.1	1119	27.9	443	26.7
Other or unknown	61	1.0	31	0.8	7	0.4
Sexual preference						

Homosexual	2260	37.5	1563	38.9	754	45.4
Bisexual	338	5.6	236	5.9	109	6.6
Heterosexual	2260	37.5	2199	54.8	793	47.7
Unknown	36	0.6	15	0.4	6	0.4
Sexual contacts						
No partner	1042	17.3	647	16.1	232	14.0
Unprotected sex with stable partner	1185	19.7	796	19.8	307	18.5
Protected sex with stable partner	1135	18.8	695	17.3	278	16.7
Unprotected sex with occasional partner	1089	18.1	822	20.5	389	23.4
Protected sex with occasional partner	985	16.3	687	17.1	302	18.2
Unknown	591	9.8	366	9.1	153	9.2
History of injection drug use						
Never	5149	85.4	3445	85.8	1354	81.5
Former	646	10.7	380	9.5	204	12.3
Current	232	3.8	188	4.7	104	6.3

[a]Value represented as median (interquartile range)

Abbreviations: SHCS Swiss HIV Cohort Study

Table 2. Pearson Correlation Coefficients Between Pairs of Socio-Behavioural Groups Across Swiss Transmission Clusters From all Subtypes, B Subtype Only and non-B Subtypes, SHCS, 2000-2011.

Socio-behavioural group	1) Gay and bisexual men of lower SEP	2) Gay and heterosexual men of higher SEP	3) Gay migrant men	4) Heterosexual people of lower SEP	5) Older heterosexual and gay people on welfare benefits	6) Migrant women in heterosexual partnerships	7) Heterosexual migrants on welfare benefits
All subtypes							
2) Gay and heterosexual men of higher SEP	0.77 [a]						
3) Gay migrant men	0.65 [a]	0.59 [a]					
4) Heterosexual people of lower SEP	0.17	0.49 [a]	0.17				
5) Older heterosexual and gay people on welfare benefits	0.48 [a]	0.67 [a]	0.34 [a]	0.75 [a]			
6) Migrant women in heterosexual partnerships	-0.07	0.08	0.12	0.43 [a]	0.28		
7) Heterosexual migrants on welfare benefits	-0.02	0.07	0.04	0.30	0.24	0.35	
8) Injection drug users	0.11	0.45 [a]	0.11	0.91 [a]	0.65 [a]	0.34	0.24

B subtype							
2) Gay and heterosexual men of higher SEP	0.76 [a]						
3) Gay migrant men	0.68 [a]	0.62 [a]					
4) Heterosexual people of lower SEP	0.18	0.51 [a]	0.17				
5) Older heterosexual and gay people on welfare benefits	0.52 [a]	0.71 [a]	0.36	0.77 [a]			
6) Migrant women in heterosexual partnerships	0.03	0.33	0.04	0.75 [a]	0.49 [a]		
7) Heterosexual migrants on welfare benefits	0.21	0.40	0.26	0.67 [a]	0.53 [a]	0.61 [a]	
8) Injection drug users	0.09	0.45 [a]	0.11	0.93 [a]	0.69 [a]	0.75 [a]	0.67 [a]
Non B subtypes							
2) Gay and heterosexual men of higher SEP	0.47 [a]						
3) Gay migrant men	0.23	0.15					
4) Heterosexual people of lower SEP	-0.14	0.15	0.14				
5) Older heterosexual and gay people on welfare benefits	0.02	0.34	0.12	0.49 [a]			
6) Migrant women in heterosexual partnerships	-0.1	-0.08	0.53 [a]	0.47 [a]	0.22		
7) Heterosexual migrants on welfare benefits	-0.17	-0.05	-0.04	0.19	0.15	0.13	
8) Injection drug users	-0.15	-0.13	-0.12	0.04	-0.12	-0.13	-0.13

[a]Value >0.4

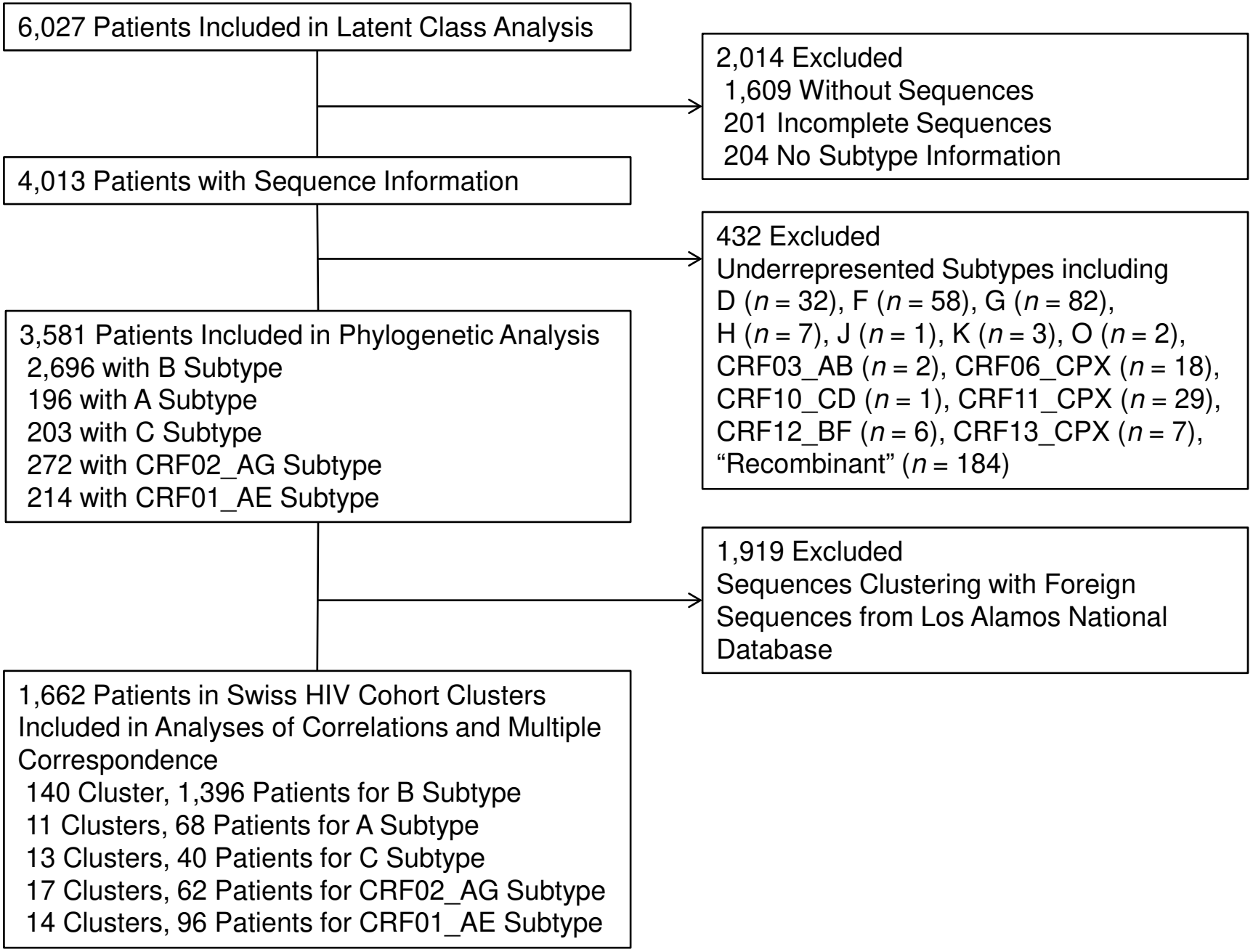
Abbreviations: SHCS Swiss HIV Cohort Study, SES socioeconomic status

Figure Legends

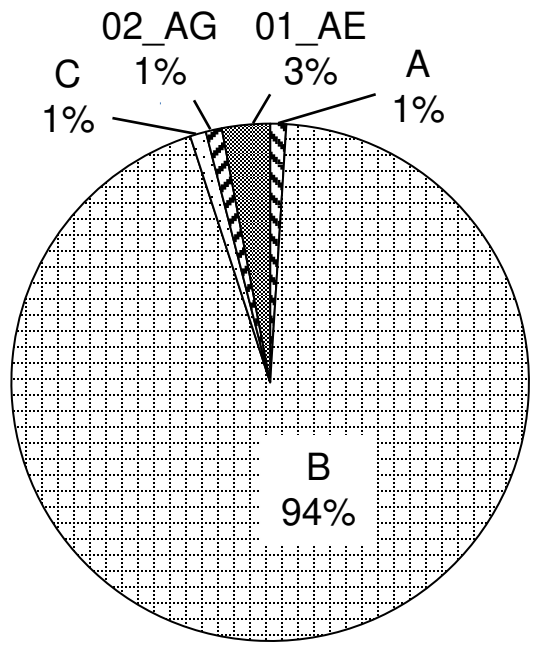
Figure 1. Selection of study populations. The flowchart depicts the number of Swiss HIV Cohort Study (SHCS) patients included in the Latent Class Analysis (LCA), phylogenetic analysis and cluster analysis. SHCS, 2000-2011.

Figure 2. Distribution of HIV viral subtypes across socio-behavioural groups in the Swiss HIV Cohort Study (SHCS). A) Gay and bisexual men of lower socioeconomic position (SEP), B) Gay and heterosexual men of higher SEP, C) Young gay migrant men, D) Heterosexual people of lower SEP, E) Older heterosexual and gay people on welfare benefits, F) Migrant women in heterosexual partnerships, G) Heterosexual migrants on welfare benefits, H) Injection drug users. SHCS, 2000-2011.

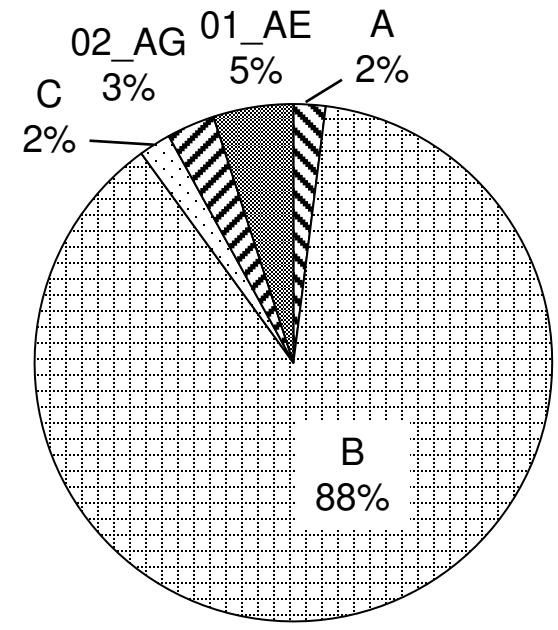
Figure 3. Concordance between socio-behavioural groups and phylogenetic clusters: A) All subtypes, B) Subtype B, C) Non-B subtypes. The proximity between socio-behavioural groups (circles) reflects their tendency to co-occur in the same phylogenetic clusters, while phylogenetic clusters (squares) are positioned in the proximity of the groups from which they contain the most sequences. Socio-behavioral groups are numbered as: 1 Gay and bisexual men of lower socioeconomic position (SEP), 2 Gay and heterosexual men of higher SEP, 3 Young gay migrant men, 4 Heterosexual people of lower SEP, 5 Older heterosexual and gay people on welfare benefits, 6 Migrant women in heterosexual partnerships, 7 Heterosexual migrants on welfare benefits, 8 Injection drug users. The two main axes explaining most of the associations between socio-behavioural groups and phylogenetic clusters as measured by the χ^2-metric (percentage in parenthesis) are shown. Results from multiple correspondence analysis, Swiss HIV Cohort Study, 2000-2011.


6,027 Patients Included in Latent Class Analysis
2,014 Excluded
1,609 Without Sequences
201 Incomplete Sequences
204 No Subtype Information
4,013 Patients with Sequence Information
432 Excluded
Underrepresented Subtypes including
D (n = 32), F (n = 58), G (n = 82),
H (n = 7), J (n = 1), K (n = 3), O (n = 2),
CRF03_AB (n = 2), CRF06_CPX (n = 18),
CRF10_CD (n = 1), CRF11_CPX (n = 29),
CRF12_BF (n = 6), CRF13_CPX (n = 7),
"Recombinant" (n = 184)
3,581 Patients Included in Phylogenetic Analysis
2,696 with B Subtype
196 with A Subtype
203 with C Subtype
272 with CRF02_AG Subtype
214 with CRF01_AE Subtype
1,919 Excluded
Sequences Clustering with Foreign
Sequences from Los Alamos National
Database
1,662 Patients in Swiss HIV Cohort Clusters
Included in Analyses of Correlations and Multiple
Correspondence
140 Cluster, 1,396 Patients for B Subtype
11 Clusters, 68 Patients for A Subtype
13 Clusters, 40 Patients for C Subtype
17 Clusters, 62 Patients for CRF02_AG Subtype
14 Clusters, 96 Patients for CRF01_AE Subtype

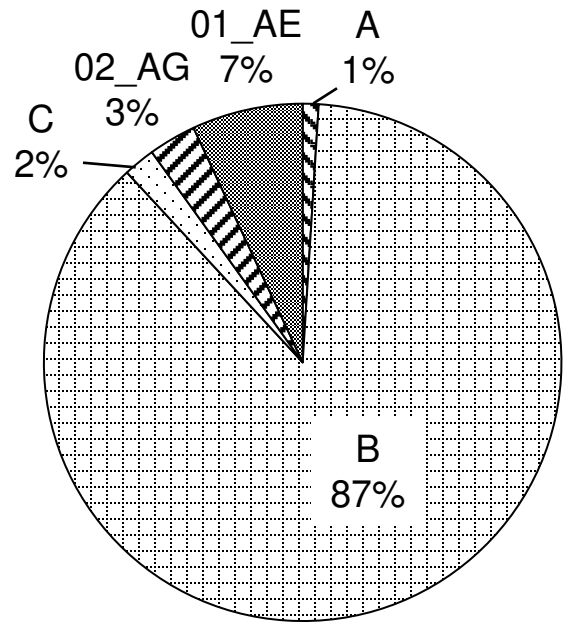
A)


02_AG
1%
01_AE
3%
C
1%
A
1%
B
94%

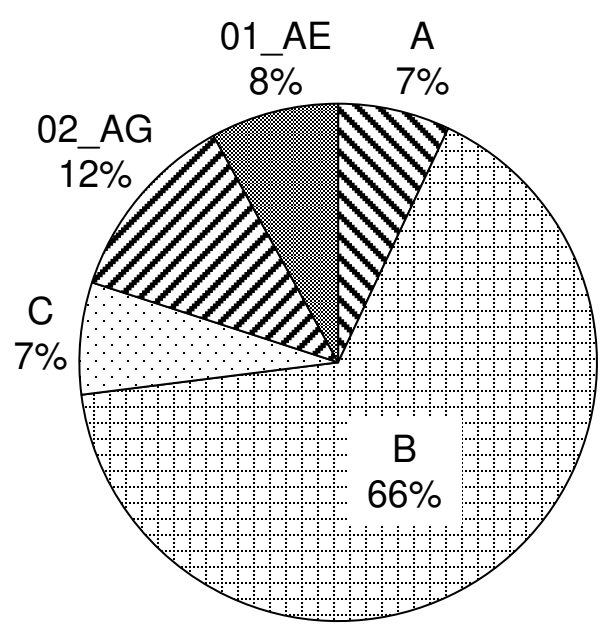
B)


02_AG
3%
01_AE
5%
A
2%
C
2%
B
88%

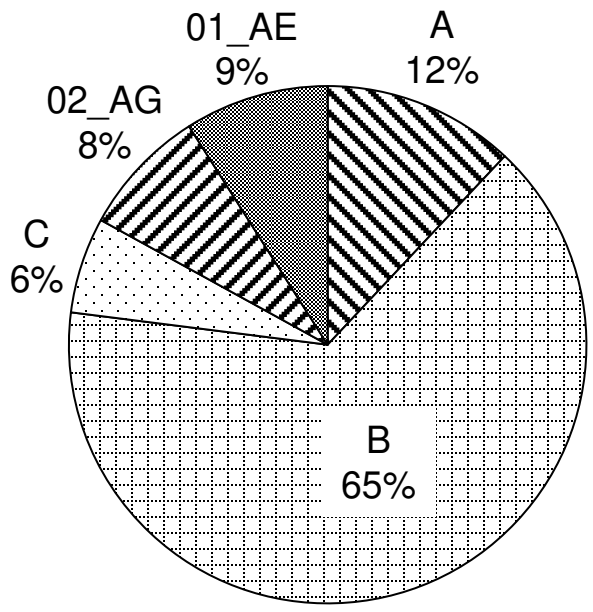
C)


01_AE
7%
A
1%
02_AG
3%
C
2%
B
87%

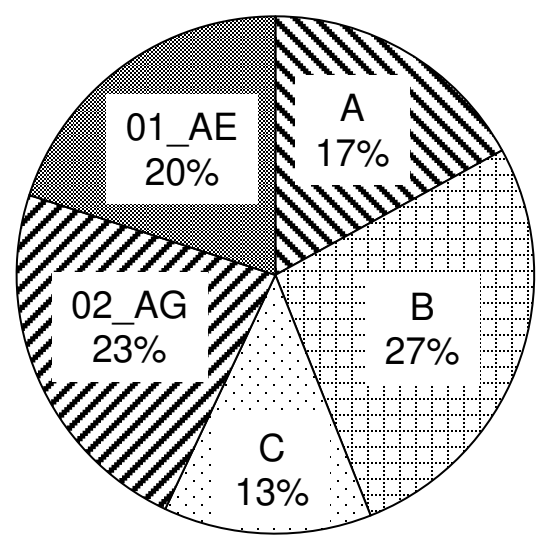
D)


01_AE
8%
A
7%
02_AG
12%
C
7%
B
66%

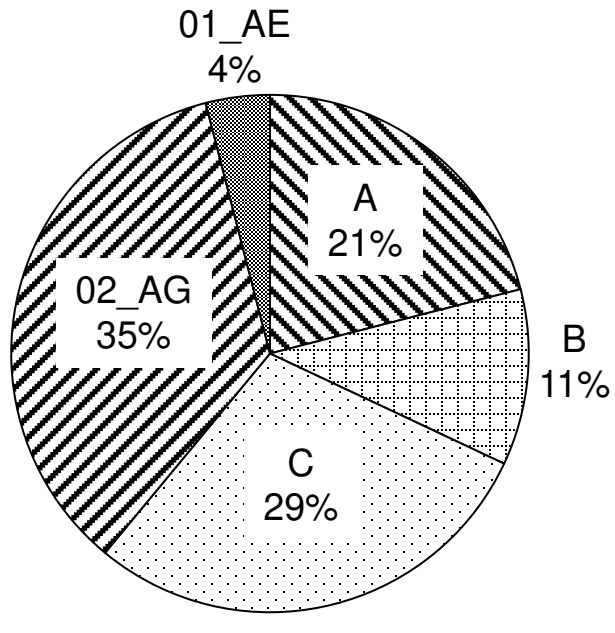
E)


01_AE
9%
A
12%
02_AG
8%
C
6%
B
65%

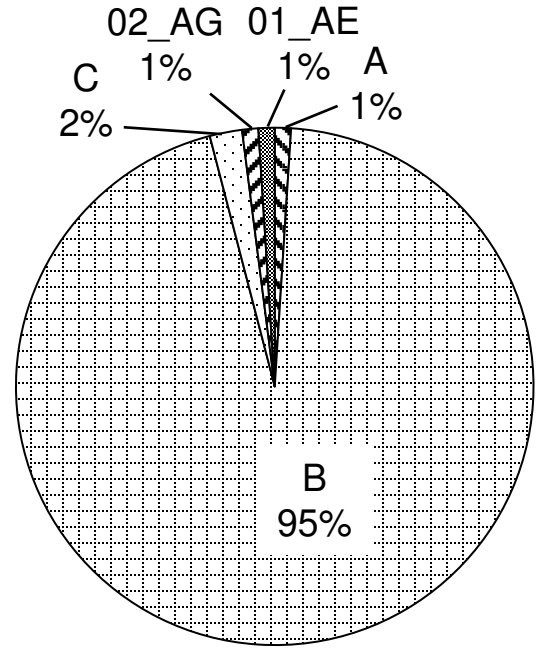
F)

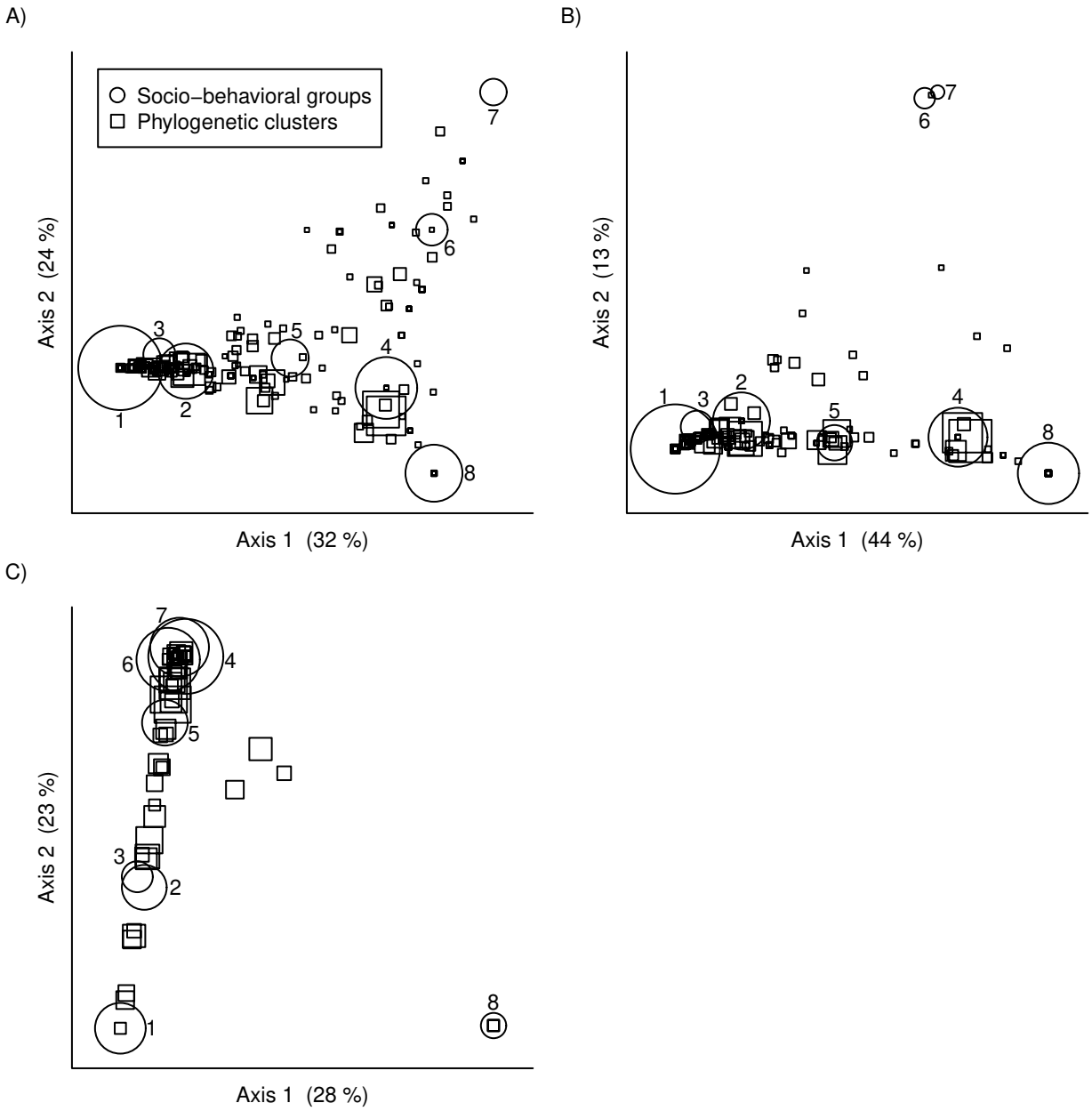

01_AE
20%
A
17%
02_AG
23%
B
27%
C
13%

G)


01_AE
4%
A
21%
02_AG
35%
B
11%
C
29%

H)


02_AG
1%
01_AE
1%
C
2%
A
1%
B
95%


A)
Socio−behavioral groups
Phylogenetic clusters
Axis 2 (24 %)
Axis 1 (32 %)
B)
Axis 2 (13 %)
Axis 1 (44 %)
C)
Axis 2 (23 %)
Axis 1 (28 %)

Social Meets Molecular: Combining Phylogenetic and Latent Class Analyses to Understand HIV-1 Transmission in Switzerland

Web Materials

Web Appendix 1: Identifying socio-behavioral groups

We used latent class analysis (1, 2) to identify groups of patients with similar socio-demographic and behavioral characteristics. We included all individuals who enrolled in the SHCS between 2000 and 2011.

Specification of latent class model

For a set of q categorical manifest categorical variables Y_l with c_l categories $(l = 1,...,q)$, a latent class model with g latent classes can be specified as follows:

$$P(Y = y) = \sum_{c=1}^{g} \pi_c \prod_{l=1}^{q} \varphi_{cly_l},$$

where Y is a vector containing the elements Y_l, π_c represents the prevalence of class c and φ_{cly_l} represents the probability that Y_l takes on the category y_l given class c ($c = 1,...,g$, $l = 1,...,q$ and $y_l = 1,...,c_l$). Thus within classes, the variables are assumed to be independent (assumption of local independence) and distributed according to the multinomial probabilities φ_{cly_l}.

We included the following 10 categorical variables

1. Sex (male, female)
2. Age (<25, 25-34, 35-44, 45-54, ≥55 years)
3. Region of origin (Switzerland and North-West Europe, Southern Europe, Sub-Saharan Africa, Latin America, Asia and Eastern Europe)
4. Level of education (compulsory schooling, vocational training, higher education)
5. Occupation (self-employed, apprentice or trainee, higher management, middle or lower management level, employee, housewife/homemaker)

6. Main source of income (salaried work, support from family or partner, welfare benefits)
7. Sexual preference (heterosexual, bisexual, homosexual)
8. Sexual contacts (no partner, unprotected sex with stable partner, protected sex with stable partner, unprotected sex with occasional partner, protected sex with occasional partner)
9. history of injection drug use (never, ever, current)
10. alcohol consumption (severe, moderate, light according to World Health Organization (3))

Variable 8 is a combination of responses to two questions, one on condom use (always, sometimes, never, no answer) and one on type of partners (stable, occasional). Sexual contacts were defined as unprotected if condom use was reported as sometimes or never. For all questions, response categories "no answer" or 'other' were recoded to missing values.

Fitting the latent class model

Models were fitted using the Mplus software version 6.1 (4) with the number of classes g varying between 1 and 10. Model fitting was based on maximum likelihood using the expectation maximization (EM) algorithm (5). We did not exclude patients with missing values in any of the variables. The EM algorithm allows to incorporate the available data from these patients under the assumption that missing values are missing at random (6).

We used the Bayesian Information Criterion to select the number of classes, thus balancing parsimony and model fit (2). We also report the Akaike Information Criterion and the entropy (7) for each model.

For each patient, we computed posterior probabilities of belonging to the different latent classes of a fitted model. We allocated individuals to the groups for which they had the highest posterior membership probability (2).

Web Table 1: Characteristics of Socio-Behavioral Groups Identified by Latent Class Analysis.

Number of participants (%) is shown unless otherwise indicated. Analysis based on 6027 participants enrolled in the Swiss HIV Cohort Study since 1 January 2000.

	1) Gay and bisexual men of lower SEP	2) Gay and heterosexual men of higher SEP	3) Young gay migrant men	4) Heterosexual people of lower SEP	5) Older heterosexual and gay people on welfare benefits	6) Migrant women in partnerships	7) Heterosexual migrants on welfare benefits	8) Injection drug users
No. of participants	1478	873	316	1089	443	546	578	704
No. of women	0 (0)	1 (0.1)	4 (1.3)	514 (47.2)	107 (24.2)	537 (98.4)	355 (61.4)	248 (35.2)
Median (interquartile range) age, years	37 (31-42)	43 (37-49)	30 (26-34)	39 (33-44)	59 (52-64)	31 (27-36)	32 (25-37)	36 (31-41)
Region of origin								
CH and North-West Europe	1260 (85.3)	743 (85.1)	54 (17.1)	678 (62.3)	397 (89.6)	66 (12.1)	6 (1.0)	557 (79.1)
Southern Europe	140 (9.5)	23 (2.6)	27 (8.5)	132 (12.1)	40 (9.0)	11 (2.0)	0 (0)	88 (12.5)
Sub-Saharan Africa	0 (0)	11 (1.3)	19 (6.0)	137 (12.6)	2 (0.5)	273 (50.0)	529 (91.5)	6 (0.9)
Latin America	40 (2.7)	25 (2.9)	102 (32.3)	31 (2.9)	0 (0)	42 (7.7)	5 (0.9)	5 (0.7)
Asia and Eastern Europe	14 (1.0)	20 (2.3)	98 (31.0)	71 (6.5)	1 (0.2)	129 (23.6)	24 (4.2)	35 (5.0)
Other or unknown	24 (1.6)	51 (5.8)	16 (5.1)	40 (3.7)	3 (0.7)	25 (4.6)	14 (2.4)	13 (1.9)
Level of education								
Compulsory schooling	87 (5.9)	1 (0.1)	131 (41.5)	233 (21.4)	95 (21.4)	325 (59.6)	390 (67.6)	349 (49.6)
Vocational training	1132 (76.7)	24 (2.8)	81 (25.6)	658 (60.4)	263 (59.4)	147 (27.0)	133 (23.1)	328 (46.6)
Higher education	256 (17.3)	847 (97.0)	102 (32.3)	192 (17.6)	84 (19.0)	71 (13.0)	47 (8.2)	22 (3.1)
Other or unknown	1 (0.1)	1 (0.1)	2 (0.6)	6 (0.6)	1 (0.2)	2 (0.4)	7 (1.2)	5 (0.7)
Occupation								
Student or trainee	8 (0.5)	0 (0)	54 (17.1)	0 (0)	0 (0)	40 (7.3)	25 (4.3)	35 (5.0)
Employee*	771 (52.2)	0 (0)	106 (33.5)	634 (58.2)	186 (42.0)	151 (27.7)	250 (43.3)	471 (66.9)
Middle or lower management	130 (8.8)	238 (27.3)	2 (0.6)	75 (6.9)	29 (6.6)	12 (2.2)	15 (2.6)	19 (2.7)
Higher management	0 (0)	186 (21.3)	1 (0.3)	15 (1.4)	7 (1.6)	0 (0)	0 (0)	0 (0)
Self-employed	131 (8.9)	149 (17.1)	24 (7.6)	122 (11.2)	70 (15.8)	38 (7.0)	48 (8.3)	17 (2.4)
Houseman or housewife	0 (0)	0 (0)	7 (2.2)	0 (0)	8 (1.8)	160 (29.3)	38 (6.6)	17 (2.4)

Other or unknown	438 (29.6)	300 (34.4)	122 (38.6)	243 (22.3)	143 (32.3)	145 (26.6)	202 (35.0)	145 (20.6)
Main source of income								
Salaried work	1312 (88.8)	796 (91.2)	152 (48.1)	963 (88.4)	112 (25.3)	46 (8.4)	91 (15.7)	82 (11.7)
Support from family or partner	0 (0)	3 (0.3)	136 (43.0)	48 (4.4)	0 (0)	432 (79.1)	18 (3.1)	19 (2.7)
Welfare benefits	163 (11.0)	71 (8.1)	21 (6.7)	72 (6.6)	328 (74.0)	59 (10.8)	450 (77.9)	592 (84.1)
Other or unknown	3 (0.2)	3 (0.3)	7 (2.2)	6 (0.6)	3 (0.7)	9 (1.7)	19 (3.3)	11 (1.6)
Sexual preference								
Homosexual	1262 (85.4)	640 (73.3)	261 (82.6)	0 (0.0)	72 (16.3)	4 (0.7)	4 (0.7)	17 (2.4)
Bisexual	164 (11.1)	78 (8.9)	27 (8.5)	1 (0.1)	29 (6.6)	2 (0.4)	10 (1.7)	27 (3.8)
Heterosexual	45 (3.0)	148 (17.0)	25 (7.9)	1082 (99.4)	335 (75.6)	537 (98.4)	563 (97.4)	658 (93.5)
Unknown	7 (0.5)	7 (0.8)	3 (1.0)	6 (0.6)	7 (1.6)	3 (98.4)	1 (0.2)	2 (0.3)
Sexual contacts								
No partner	107 (7.2)	82 (9.4)	4 (1.3)	116 (10.7)	189 (42.7)	9 (1.7)	273 (47.2)	262 (37.2)
Unprotected sex with stable partner	148 (10.0)	105 (12.0)	37 (11.7)	412 (37.8)	47 (10.6)	264 (48.4)	83 (14.4)	89 (12.6)
Protected sex with stable partner	133 (9.0)	135 (15.5)	117 (37.0)	300 (27.6)	47 (10.6)	210 (38.5)	70 (12.1)	123 (17.5)
Unprotected sex with occasional partner	503 (34.0)	252 (28.9)	65 (20.6)	98 (9.0)	51 (11.5)	9 (1.7)	49 (8.5)	62 (8.8)
Protected sex with occasional partner	490 (33.2)	216 (24.7)	73 (23.1)	40 (3.7)	28 (6.3)	4 (0.7)	53 (9.2)	81 (11.5)
Unknown	97 (6.6)	83 (9.5)	20 (6.3)	123 (11.3)	81 (18.3)	50 (9.2)	50 (8.7)	87 (12.4)
History of infection drug use								
Never	1438 (97.3)	844 (96.7)	307 (97.2)	961 (88.3)	427 (96.4)	540 (98.9)	578 (100)	54 (7.7)
Ever	40 (2.7)	28 (3.2)	9 (2.9)	123 (11.3)	15 (3.4)	6 (1.1)	0 (0)	425 (60.4)
Current	0 (0)	1 (0.1)	0 (0)	5 (0.5)	1 (0.2)	0 (0)	0 (0)	225 (32.0)
Alcohol consumption**								
Severe	24 (1.6)	13 (1.5)	3 (1.0)	18 (1.7)	18 (4.1)	12 (2.2)	0 (0.0)	53 (7.5)
Moderate	49 (3.3)	41 (4.7)	9 (2.9)	69 (6.3)	24 (5.4)	11 (2.0)	11 (1.9)	46 (6.5)
Light	1310 (88.6)	748 (85.7)	278 (88.0)	902 (82.8)	356 (80.4)	478 (87.6)	497 (86.0)	425 (60.4)
Unknown	95 (6.4)	71 (8.1)	26 (8.2)	100 (9.2)	45 (10.2)	45 (8.2)	70 (12.1)	180 (25.6)

Abbreviations: CH Switzerland, SEP socio-economic position

* Manual or non-manual

** Categorization according to according to World Health Organization (3)

Web Table 2: Model diagnostics for latent class models with 1 to 10 latent classes

Number of latent classes	Log likelihood	No. of free Parameters	Akaike Information Criterion	Bayesian Information Criterion	Entropy *
1	-54296.00	28	108648	108836	
2	-50179.60	57	100473	100855	0.844
3	-49089.46	86	98351	98927	0.831
4	-48575.71	115	97381	98152	0.867
5	-48228.89	144	96746	97711	0.824
6	-47925.15	173	96196	97356	0.821
7	-47674.36	202	95753	97107	0.806
8	-47457.23	231	95376	96925	0.789
9	-47342.85	260	95206	96949	0.787
10	-47282.90	289	95144	97081	0.800

* This entropy measure ranges from 0 to 1, with 0 indicating highly uncertain membership of subjects to latent classes (equal membership probabilities for all latent classes) and 1 indicating certain membership (membership probabilities are either 1 or 0) (7).

Web Table 3: Distribution of socio-behavioral variables (probabilities of response categories) within latent classes as estimated by the selected model with 8 classes

Latent class	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
Class prevalence*	0.251	0.132	0.060	0.169	0.081	0.096	0.094	0.117
Mean posterior probabilities**	0.862	0.767	0.838	0.766	0.815	0.879	0.832	0.892
Label assigned	Gay and bisexual men of lower SEP	Gay and heterosexual men of higher SEP	Young gay migrant men	Heterosexual people of lower SEP	Older heterosexual and gay people on welfare	Migrant women in heterosexual partnerships	Heterosexual migrants on welfare benefits	Injection drug users
Sex								
Male	1.000	0.997	0.984	0.528	0.761	0.021	0.386	0.653
Female	0.000	0.003	0.016	0.472	0.239	0.979	0.614	0.347
Age								
15-24 years	0.042	0.000	0.165	0.034	0.000	0.118	0.210	0.051
25-34 years	0.331	0.145	0.547	0.262	0.000	0.569	0.434	0.371
35-44 years	0.446	0.454	0.238	0.469	0.091	0.253	0.286	0.470
45-54 years	0.155	0.273	0.050	0.203	0.302	0.046	0.053	0.108
≥55 years	0.025	0.129	0.000	0.031	0.607	0.013	0.016	0.000
Region of origin								
CH and North-West Europe	0.869	0.895	0.261	0.640	0.892	0.140	0.028	0.804
Southern Europe	0.087	0.033	0.114	0.121	0.096	0.022	0.000	0.129
Sub-Saharan Africa	0.002	0.018	0.051	0.142	0.007	0.533	0.909	0.008
Latin America	0.029	0.029	0.294	0.029	0.000	0.077	0.013	0.007
Asia and Eastern Europe	0.012	0.026	0.280	0.067	0.006	0.228	0.050	0.051
Level of education								
Compulsory schooling	0.055	0.003	0.390	0.223	0.207	0.577	0.683	0.491
Vocational training	0.707	0.103	0.291	0.583	0.583	0.296	0.231	0.480
Higher education	0.237	0.894	0.320	0.193	0.210	0.126	0.086	0.030
Occupation								
Self-employed	0.127	0.260	0.123	0.141	0.240	0.091	0.123	0.041
Apprentice or trainee	0.009	0.000	0.232	0.000	0.000	0.091	0.070	0.062
Higher management	0.004	0.334	0.006	0.024	0.022	0.000	0.000	0.000
Middle or lower management	0.138	0.406	0.017	0.094	0.100	0.035	0.035	0.035
Employee†	0.722	0.000	0.591	0.741	0.616	0.412	0.662	0.832

Houseman or housewife	0.000	0.000	0.030	0.000	0.022	0.371	0.110	0.030
Main source of income								
Salaried work	0.894	0.907	0.544	0.849	0.333	0.110	0.202	0.143
Support from family or partner	0.000	0.007	0.370	0.050	0.002	0.745	0.042	0.027
Welfare benefits	0.106	0.086	0.087	0.101	0.665	0.145	0.756	0.830
Sexual preference								
Homosexual	0.850	0.732	0.817	0.000	0.178	0.007	0.006	0.024
Bisexual	0.103	0.091	0.096	0.003	0.070	0.003	0.017	0.039
Heterosexual	0.046	0.177	0.087	0.997	0.752	0.990	0.977	0.937
Sexual contacts								
No partner	0.082	0.105	0.021	0.137	0.475	0.028	0.504	0.410
Unprotected sex with stable partner	0.110	0.140	0.122	0.405	0.153	0.527	0.166	0.155
Protected sex with stable partner	0.098	0.167	0.368	0.308	0.136	0.417	0.139	0.209
Unprotected sex with occasional partner	0.364	0.310	0.238	0.098	0.157	0.020	0.095	0.097
Protected sex with occasional partner	0.346	0.279	0.250	0.052	0.078	0.008	0.096	0.129
History of injection drug use								
Never	0.971	0.968	0.967	0.872	0.951	0.984	1.000	0.100
Ever	0.029	0.031	0.033	0.120	0.047	0.015	0.000	0.585
Current	0.000	0.001	0.000	0.008	0.002	0.001	0.000	0.315
Alcohol consumption‡								
Severe	0.015	0.020	0.012	0.023	0.044	0.019	0.000	0.096
Moderate	0.036	0.052	0.033	0.069	0.064	0.024	0.021	0.086
Light	0.949	0.928	0.956	0.908	0.892	0.957	0.979	0.818

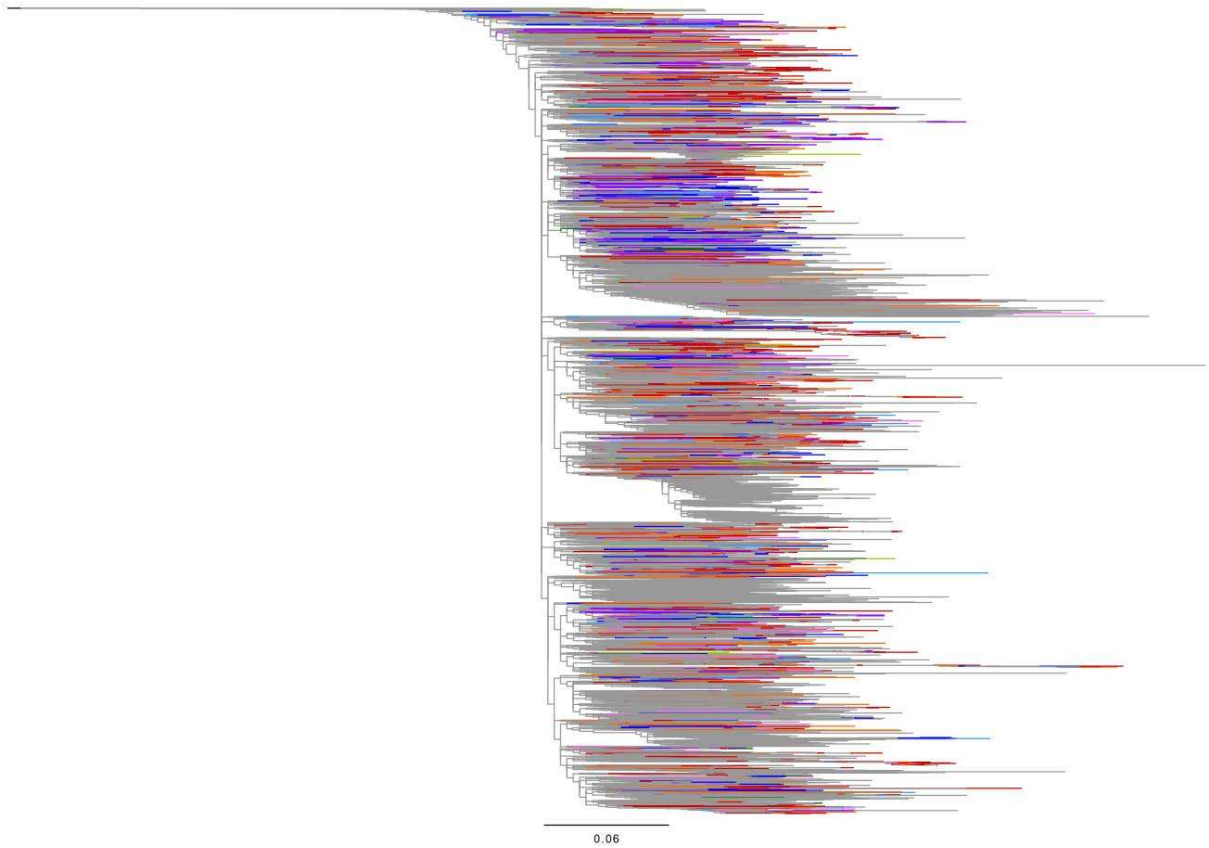
Abbreviations: CH Switzerland, SEP socio-economic position
* As estimated by the latent class model
** Among subjects whose highest membership probability is for the given class
† Manual or non-manual
‡ Categorization according to according to World Health Organization (3)

Web Table 4. Pearson Correlation Coefficients and 95% Confidence Intervals (in Parenthesis) Between Pairs of Socio-Behavioural Groups Across Swiss Transmission Clusters From all Subtypes, B Subtype Only and non-B Subtypes.

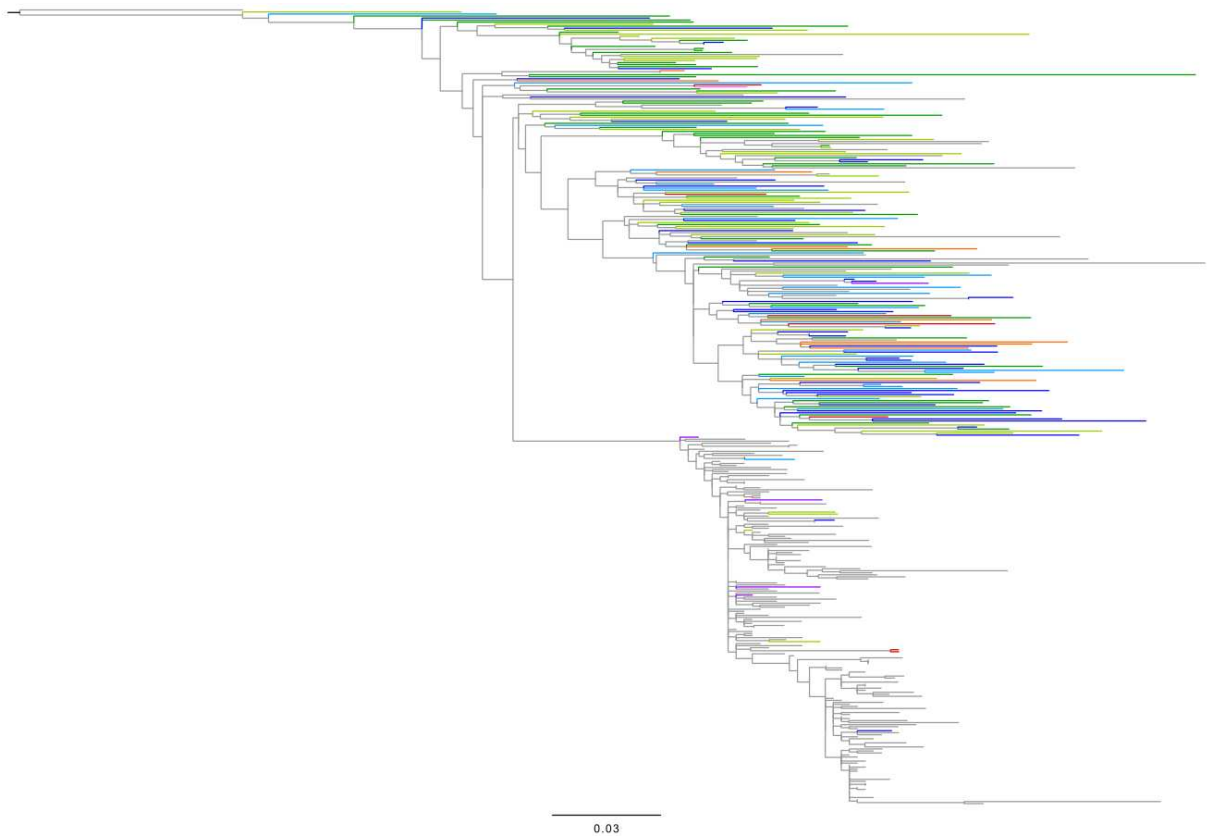
Socio-behavioural group	1) Gay and bisexual men of lower SEP	2) Gay and heterosexual men of higher SEP	3) Gay migrant men	4) Heterosexual people of lower SEP	5) Older heterosexual and gay people on welfare benefits	6) Migrant women in heterosexual partnerships	7) Heterosexual migrants on welfare benefits
All subtypes							
2) Gay and heterosexual men of higher SEP	0.77 (0.70-0.82)						
3) Gay migrant men	0.65 (0.56-0.72)	0.59 (0.49-0.68)					
4) Heterosexual people of lower SEP	0.17 (0.03-0.30)	0.49 (0.38-0.59)	0.17 (0.03-0.30)				
5) Older heterosexual and gay people on welfare benefits	0.48 (0.36-0.58)	0.67 (0.58-0.74)	0.34 (0.21-0.46)	0.75 (0.68-0.80)			
6) Migrant women in heterosexual partnerships	-0.07 (-0.21-0.07)	0.08 (-0.06-0.21)	0.12 (-0.02-0.25)	0.43 (0.31-0.54)	0.28 (0.15-0.41)		
7) Heterosexual migrants on welfare benefits	-0.02 (-0.16-0.12)	0.07 (-0.07-0.21)	0.04 (-0.10-0.18)	0.30 (0.17-0.42)	0.24 (0.11-0.37)	0.35 (0.22-0.47)	
8) Injection drug users	0.11 (-0.03-0.25)	0.45 (0.33-0.55)	0.11 (-0.03-0.25)	0.91 (0.88-0.93)	0.65 (0.56-0.72)	0.34 (0.21-0.46)	0.24 (0.11-0.37)
B subtype							
2) Gay and heterosexual men of higher SEP	0.76 (0.68-0.82)						
3) Gay migrant men	0.68 (0.58-0.76)	0.62 (0.51-0.71)					
4) Heterosexual people of lower SEP	0.18 (0.01-0.33)	0.51 (0.38-0.62)	0.17 (0.00-0.33)				

5) Older heterosexual and gay people on welfare benefits	0.52 (0.39-0.63)	0.71 (0.61-0.78)	0.36 (0.21-0.50)	0.77 (0.70-0.83)			
6) Migrant women in heterosexual partnerships	0.03 (-0.13-0.20)	0.33 (0.17-0.47)	0.04 (-0.12-0.21)	0.75 (0.66-0.81)	0.49 (0.35-0.61)		
7) Heterosexual migrants on welfare benefits	0.21 (0.05-0.37)	0.40 (0.25-0.53)	0.26 (0.10-0.41)	0.67 (0.56-0.75)	0.53 (0.40-0.64)	0.61 (0.49-0.70)	
8) Injection drug users	0.09 (-0.07-0.25)	0.45 (0.30-0.57)	0.11 (-0.06-0.27)	0.93 (0.90-0.95)	0.69 (0.59-0.77)	0.75 (0.67-0.81)	0.67 (0.57-0.75)
Non B subtypes							
2) Gay and heterosexual men of higher SEP	0.47 (0.24-0.66)						
3) Gay migrant men	0.23 (-0.04-0.46)	0.15 (-0.12-0.40)					
4) Heterosexual people of lower SEP	-0.14 (-0.39-0.13)	0.15 (-0.12-0.40)	0.14 (-0.13-0.39)				
5) Older heterosexual and gay people on welfare benefits	0.02 (-0.25-0.28)	0.34 (0.08-0.55)	0.12 (-0.15-0.38)	0.49 (0.26-0.67)			
6) Migrant women in heterosexual partnerships	-0.10 (-0.35-0.17)	-0.08 (-0.34-0.19)	0.53 (0.30-0.69)	0.47 (0.23-0.65)	0.22 (-0.05-0.46)		
7) Heterosexual migrants on welfare benefits	-0.17 (-0.42-0.10)	-0.05 (-0.31-0.22)	-0.04 (-0.30-0.23)	0.19 (-0.08-0.43)	0.15 (-0.12-0.40)	0.13 (-0.14-0.38)	
8) Injection drug users	-0.15 (-0.40-0.12)	-0.13 (-0.38-0.14)	-0.12 (-0.37-0.15)	0.04 (-0.23-0.30)	-0.12 (-0.37-0.15)	-0.13 (-0.39-0.14)	-0.13 (-0.38-0.14)

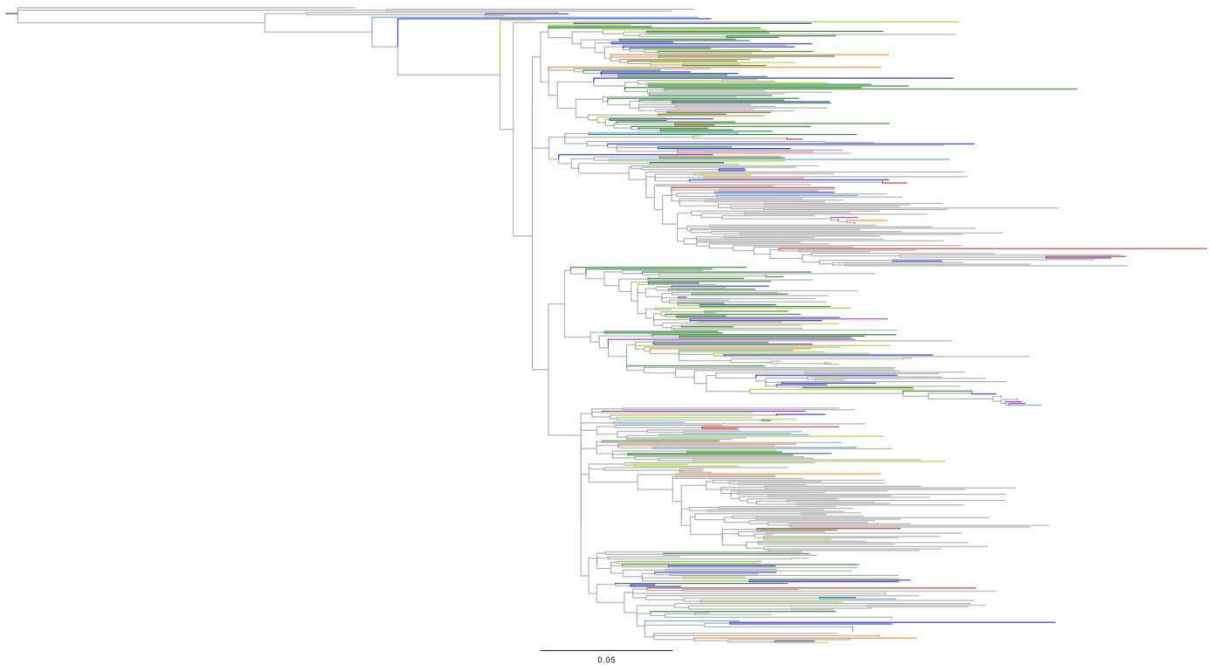
Web Figure 1: Maximum likelihood tree for subtype B, with 2697 sequences from the Swiss HIV Cohort Study (SHCS) and 2697 from Los Alamos HIV Database (LAHD). A subtype K sequence was used as an outgroup. LAHD sequences are coloured in grey, whereas SHCS sequences coloured by socio-behavioural groups identified by Latent Class Analysis: *red* for 'gay and bisexual men of lower SEP', *orange* for 'gay and heterosexual men of higher SEP', *pink* for 'young gay migrant men', *dark blue* for 'heterosexual people of lower SEP', *light blue* for 'older heterosexual and gay people on welfare benefits', *light green* for 'migrant women in partnerships', *dark green* for 'heterosexual migrants on welfare benefits' and *purple* for 'injection drug users'.



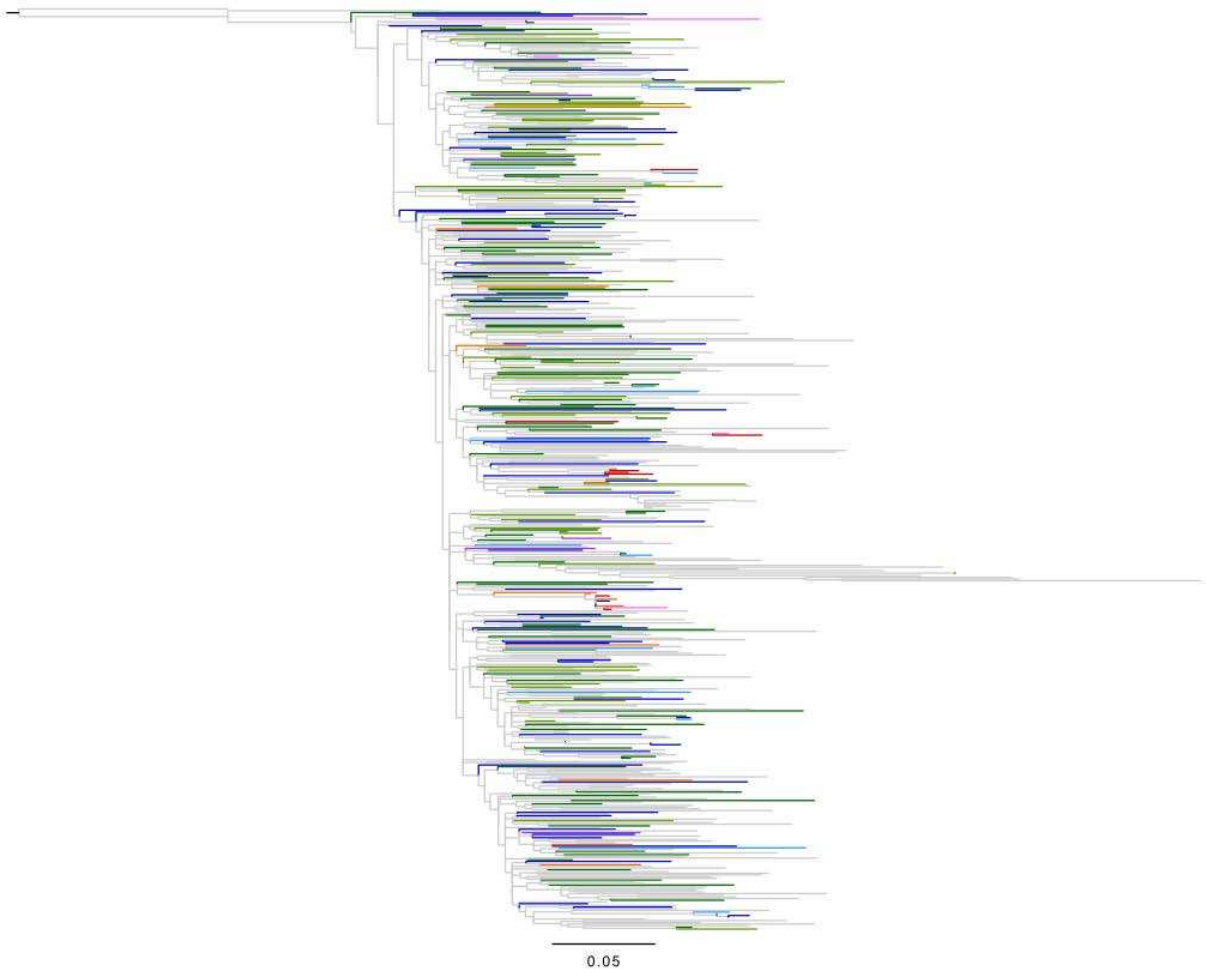
Web Figure 2: Maximum likelihood tree for subtype A, with 196 sequences from the Swiss HIV Cohort Study (SHCS) and 196 from Los Alamos HIV Database (LAHD). A subtype K sequence was used as an outgroup. LAHD sequences are coloured in grey, whereas SHCS sequences coloured by socio-behavioural groups identified by Latent Class Analysis: *red* for 'gay and bisexual men of lower SEP', *orange* for 'gay and heterosexual men of higher SEP', *pink* for 'young gay migrant men', *dark blue* for 'heterosexual people of lower SEP', *light blue* for 'older heterosexual and gay people on welfare benefits', *light green* for 'migrant women in partnerships', *dark green* for 'heterosexual migrants on welfare benefits' and *purple* for 'injection drug users'.



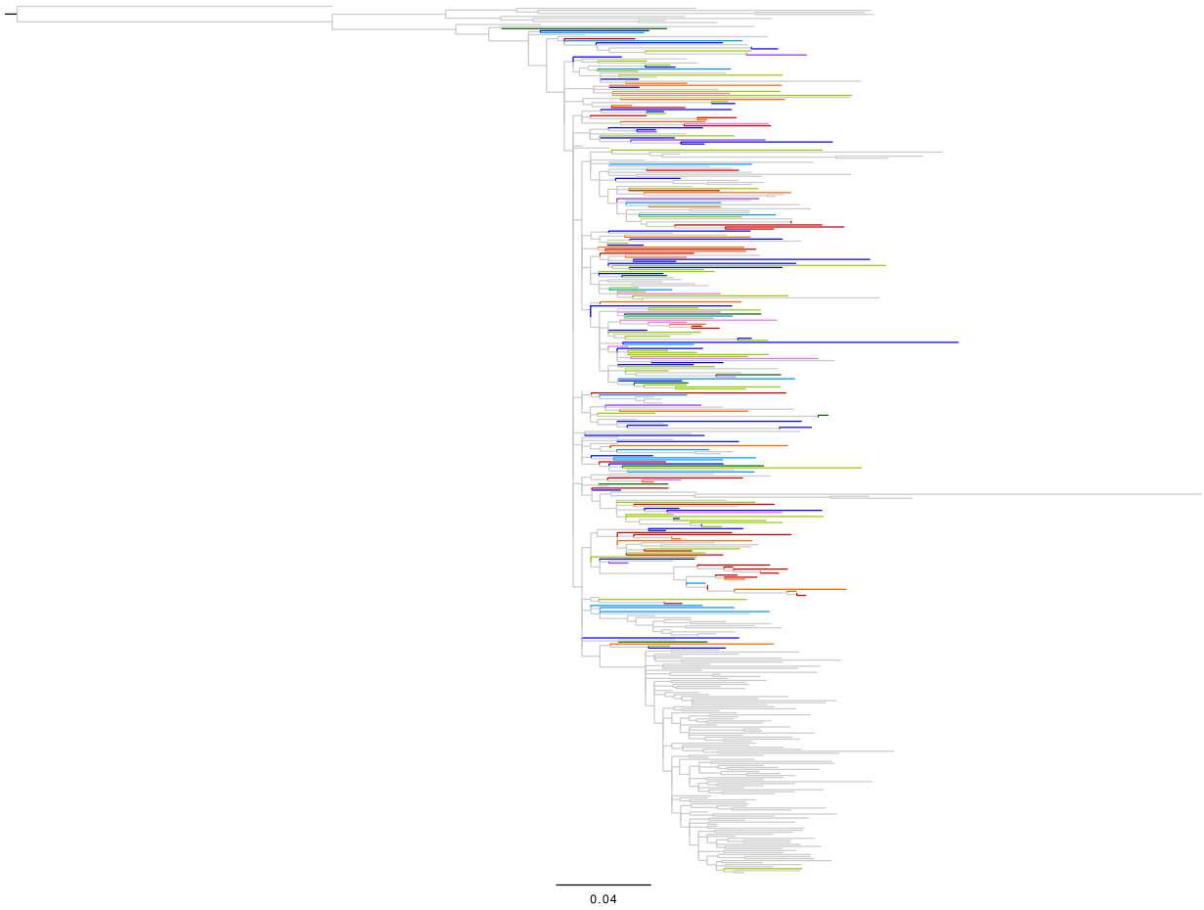
Web Figure 3: Maximum likelihood tree for subtype C, with 203 sequences from the Swiss HIV Cohort Study (SHCS) and 203 from Los Alamos HIV Database (LAHD). A subtype K sequence was used as an outgroup. LAHD sequences are coloured in grey, whereas SHCS sequences coloured by socio-behavioural groups identified by Latent Class Analysis: *red* for 'gay and bisexual men of lower SEP', *orange* for 'gay and heterosexual men of higher SEP', *pink* for 'young gay migrant men', *dark blue* for 'heterosexual people of lower SEP', *light blue* for 'older heterosexual and gay people on welfare benefits', *light green* for 'migrant women in partnerships', *dark green* for 'heterosexual migrants on welfare benefits' and *purple* for 'injection drug users'.



Web Figure 4: Maximum likelihood tree for recombinant subtype CRF02_AG, with 272 sequences from the Swiss HIV Cohort Study (SHCS) and 272 from Los Alamos HIV Database (LAHD). A subtype K sequence was used as an outgroup. LAHD sequences are coloured in grey, whereas SHCS sequences coloured by socio-behavioural groups identified by Latent Class Analysis: *red* for 'gay and bisexual men of lower SEP', *orange* for 'gay and heterosexual men of higher SEP', *pink* for 'young gay migrant men', *dark blue* for 'heterosexual people of lower SEP', *light blue* for 'older heterosexual and gay people on welfare benefits', *light green* for 'migrant women in partnerships', *dark green* for 'heterosexual migrants on welfare benefits' and *purple* for 'injection drug users'.



Web Figure 5: Maximum likelihood tree for recombinant subtype CRF01_AE, with 214 sequences from the Swiss HIV Cohort Study (SHCS) and 214 from Los Alamos HIV Database (LAHD). A subtype K sequence was used as an outgroup. LAHD sequences are coloured in grey, whereas SHCS sequences coloured by socio-behavioural groups identified by Latent Class Analysis: *red* for 'gay and bisexual men of lower SEP', *orange* for 'gay and heterosexual men of higher SEP', *pink* for 'young gay migrant men', *dark blue* for 'heterosexual people of lower SEP', *light blue* for 'older heterosexual and gay people on welfare benefits', *light green* for 'migrant women in partnerships', *dark green* for 'heterosexual migrants on welfare benefits' and *purple* for 'injection drug users'.



References

1. Lazarsfeld PF, Henry NW. *Latent structure analysis*. New York,: Houghton; 1968.
2. McLachlan G, Peel D. *Finite mixture models*. New York: John Wiley & Sons; 2000.
3. World Health Organization. Comparative quantification of health risks - Global and regional burden of disease attribution to selected major risk factors. Geneva, 2004.
4. Muthén BO, Muthén LK. *Mplus [Computer software]*. Los Angeles, CA: Muthén & Muthén; 1998.
5. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J Roy Statist Soc Ser B* 1977;39(1):1-38.
6. Little RJA, Rubin DB. *Statistical analysis with missing data*. New York; Chichester: John Wiley & Sons; 2002.
7. Ramaswamy V, Desarbo WS, Reibstein DJ, et al. An Empirical Pooling Approach for Estimating Marketing Mix Elasticities with Pims Data. *Market Sci* 1993;12(1):103-124.